

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: Ideographic Research Group Document
Title: Request to confirm three latent M-Source glyphs in the Unification Framework of Chinese Characters Encoding for the Information Systems of Macao SAR and other related issues for TCA and UTC
Source: Eiso Chan (陈永聪, Culture and Art Publishing House)
Status: Individual Contribution to IRG #59, online meeting
Action: For consideration by Macao SARG, TCA, UTC, IRG
Date: 2022-10-06

Macao SAR has released “MSCS-2020” character set and clarified the basic composition of the Unification Framework of Chinese Characters Encoding for the Information Systems of Macao SAR (abbreviated as UCIM in this document temporarily, which is not the official abbreviation used in Macao SAR now). In addition to the issues which I mentioned in [my feedback](#) on the [appendix](#) of [IRGN2456](#), I met other issues when my friend and I tried to make a font with text or display type design used for Macao SAR. These issues are related to not only one region, so it is necessary to get a consensus in IRG. The following ones mentioned here are all the latent M-Source glyphs.

1. MB1-C255 and MA-A0E8

According to the introduction in IRGN2456, UCIM inherited all from the original character set of Big5 and the character set and the encoding mapping of HKSCS-2008. The original character set and the encoding scheme of Big5 was released in 1984 by [III](#) (Institute of Information Industry), but the confirmed version as Big5-2003 was released in 2003 by [CMEX](#) (Chinese Foundation of Digitization Technology). On the other hand, there is one extended version named as Big5+/Big5E released in 1997(?). If we treat Big5+ as the extended version of the mapping of the original Big5 used for so many OSeS, there will be a pair of duplicates. Therefore, the best basic version for Big5+ is Big5-2003.

In the original Big5 specification, the glyph for 0xC255 looks like U+5F5D (𪗎) more than U+5F5E (𪗏). Please see Fig. 1.



Fig. 1 0xC255 in the original Big5 specification

Microsoft inherited the glyph to CP950 and mapped it to U+5F5D (彝). Please see Fig. 2.

C2	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4	駿 99FF	鮮 9BAE	鮫 9BAB	鮫 9BAA	鮫 9BAD	鴻 9D3B	鴿 9D3F	糞 9E9B	黏 9ECF	點 9EDE	黜 9EDC	黜 9EDD	黛 9EDB	軒 9F3E	齋 9F4B	糞 53E2
5	噲 5695	嚮 56AE	墳 58D9	壘 58D8	塙 5B38	彝 5F5D	憑 61E3	戳 6233	擴 64F4	擴 64F2	擾 64FE	攪 6506	攏 64FA	擻 64FB	擻 64P7	斷 65B7
6	曜 66DC	曠 6726	檳 6AB3	檉 6AAC	櫃 6AC3	櫃 6ABB	榨 6AB8	權 6AC2	構 6AAE	檯 6AAP	歐 6B5F	歸 6B78	殞 6BAF	瀉 7009	瀉 700B	漣 6FPE
7	濱 7006	澗 6FFA	澤 7011	瀾 700F	燠 71FB	燠 71FC	燠 71FE	燠 71FF	獵 7377	獵 7375	壁 74A7	璋 74BF	璽 7515	璽 7656	璽 7658	

Fig.2 0xC255 in MS CP950

In UniHan Database, the kBigFive property value of U+5F5D (彝) is C255, but there is no value for U+5F5E (彝). Adobe-CNS1 also inherited this mapping as a glyph collection.

However, the mapping of 0xC255 is U+5F5E (彝) in Big5-2003, because the corresponding CNS 11643 code is 1-7641, and U+5F5E (彝) is treated as the regular one (A01266) in Taiwan Province, and U+5F5D (彝) is treated as the variant (A01266-032). Please see Fig. 3.

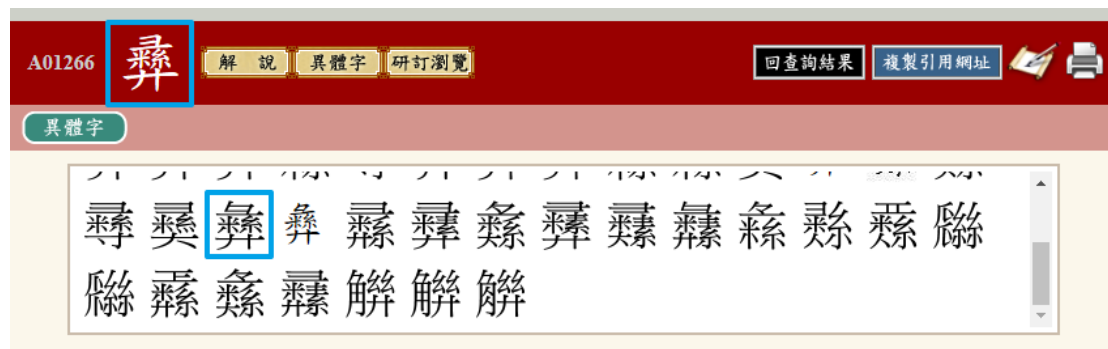


Fig. 3 A01266 in the online version of Dictionary of Chinese Variants

IBM uses the same mapping like Big5-2003. Please see Fig. 4 and 5.

220	彝	6560	A55F	T5180080
-----	---	------	------	----------

Fig. 4 T5180080 in IBM Corporate Specification C-H 3-3220-126:1999-04 IBM Traditional Chinese Graphic Character Set

Note: Column meanings are the followings, 1 for Line Number, 2 for Graphic, 3 for Host Code, 4 for PC Code, 5 for GCGID.

彝	6560	C255	5F5E	T5180080
---	------	------	------	----------

Fig. 5 T5180080 in IBM Corporate Specification C-H 3-3220-126:2016-04 IBM Traditional Chinese Graphic Character Set

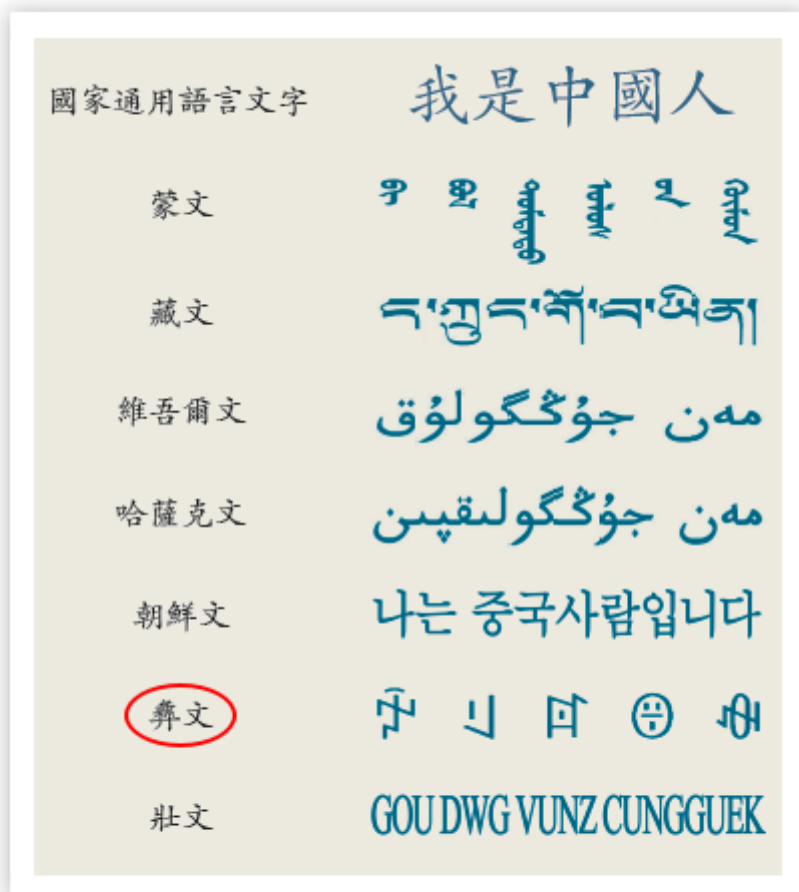
Note: Column meanings are the followings, 1 for Graphic, 2 for Host Code, 3 for Big-5 Code, 4 for UCS-2, 5 for GCGID.

On the other hand, U+5F5D (彝) is also included as 0x82EE in Big5+, which the corresponding CNS 11643 code is 3-5764.

Therefore, there are two mappings for 0xC255 now, and they are related to SCS in UCS and Unicode.

I just check the use of these two UCS characters in Macao. The textbooks all use U+5F5D (彝),

which the meaning is related to Yi people and the [Suez Canal](#). Please see Fig. 6 through 8. These textbooks are all edited or published by the [Education and Youth Development Bureau](#) of the Government of the Macao Special Administrative Region, PRC (DSEDJ) and the [People Education Press](#) (PEP), so they are authoritative for Macao SAR educations.



幾種民族文字“我是中國人”的寫法

Fig. 6 Primary School Constitutional Education Textbook, p. 49

人民教育出版社課程教材研究所，澳門特別行政區政府教育及青年發展局：《澳門教材·試行本 憲法教育 小學篇》，北京：人民教育出版社，2021年11月，ISBN 978-7-107-36197-5



2018年2月13日，“衣錦風尚——雲南省和內蒙古自治區民族服飾展”在澳門盧廉若公園春草堂揭幕。此次展覽除展出彝族、哈尼族、白族、蒙古族、達斡爾族、鄂倫春族等少數民族的服飾以外，還設置了傳統工藝演示，包括傣族織錦、白族甲馬，以及內蒙古蘇尼特服飾和烏拉特銅銀器製作等，由國家級非物質文化遺產傳承人現場示範。



充滿民族風情的少數民族服飾

Fig. 7 Secondary School Constitution Education Textbook, p. 40

人民教育出版社課程教材研究所，澳門特別行政區政府教育及青年發展局：《澳門教材試行本 憲法教育 中學篇》，北京：人民教育出版社，2021年11月，ISBN 978-7-107-36202-6

從澳門向東，經南海、太平洋可達北美洲、南美洲各國。向西，經南海、馬六甲海峽、印度洋、紅海、蘇彝士運河，進入地中海，可達地中海沿岸的南歐、西亞、北非各國。

東北方向，經台灣海峽、東海，可到日本、韓國、朝鮮及俄羅斯各港口（圖1.7）。東南方向，經南海，可到達菲律賓、印度尼西亞、澳大利亞、新西蘭及南太平洋各國的港口。

Fig. 8 Macao Geography Textbook, p. 7

人民教育出版社課程教材研究所：《澳門地理——初中地理補充教材》，澳門：澳門特別行政區政府教育及青年發展局；北京：人民教育出版社，2014年3月，ISBN 978-99965-44-30-9

I also check these two UCS characters on Macao Daily website in Sept. 29th, 2022, there are 92 results for U+5F5D (彝), only 1 result for U+5F5E (彝).

【澳門日報消息】由澳門緬甸友好協會組織的“走訪雲南僑鄉聯誼使澳門更好地融入‘一帶一路’建設”商務旅遊交流團，今日上午參觀了石林景區，感受氣勢磅礴、逶迤連綿的石海奇景。下午抵達楚雄彝人古鎮，領略博大精深的彝族文化和絢麗多姿的民族風情。

石林位於雲南省會昆明東南郊八十餘公里處，以其雄、奇、險、幽的地貌風光獨樹一幟，更被世界遺產委員會評價為世界上喀斯特地質地貌的最好範例，具有最好的自然現象和非同尋常的美學價值、科學價值。並於二〇〇七年作為“中國南方喀斯特”的重要組成部分，成功錄入了“世界自然遺產”名錄。

Fig. 9 One example for U+5F5D
《[澳緬友好協會考察雲南旅遊資源](#)》

【澳門日報消息】四川一村落昨日發生落石傷人事件，3戶民房被砸中，嚴重受損，造成至少1人死亡，3人受傷。



落實砸中民居（網圖）

四川峨邊彝族自治縣政府新聞辦官方微博公佈，四川樂山峨邊縣沙坪鎮新聲村於昨日下午約5時有滾石砸中民居，導致3人受傷，1人被埋。

Fig. 10 One example for U+5F5E

《【有片】四川落石砸中民居 至少一死三傷》

Note: “落實” in the original picture description should be “落石”.

For IICORE, M tag is marked with U+5F5D (彝), and there is no IICORE tag for U+5F5E (彝).

All in all, the most common and suitable form of MB1-C255 for Macao SAR is U+5F5D (彝), and the glyph should be as below.

彝

Fig. 11 Suggested MB1-C255 glyph for U+5F5D in UCIM

U+5F5E (彝) is also needed for the legacy use, and the corresponding MD-5F5E glyph should be as below. Hong Kong SAR has confirmed its suitable form for a long time, so there is no need to add HD-5F5E to HKSCS now.

彝

Fig. 12 Suggested MD-5F5E glyph for U+5F5E in UCIM

U+3C18 (彝) is a character related to this pair. This character has been included in HKSCS, but the right part for current H glyph (and G glyph) is U+5F5D (彝), and T glyph is U+5F5E (彝) in

the code charts.

For the current Macao SAR conventions, the glyphs rule of all the Hanzi used in Macao SAR should be like the following shows.

Table 1 Macao SAR conventions and the reference fonts

Range	Reference Font
Mentioned in MSCS-2020	Ming_MSCS
Big5 and HKSCS-2008 in BMP	PMingLiU
Big5 and HKSCS-2008 in SIP	PMingLiU-ExtB

When Macao end-users use PMingLiU to display U+3C18 (𢮑), they will get the following glyph.



Fig. 13 Glyph for U+3C18 in PMingLiU

When they use MingLiU_HKSCS to display U+3C18 (𢮑), they will get the following glyph.



Fig. 14 Glyph for U+3C18 in MingLiU_HKSCS

There are several differences between two glyphs for Fig. 13 and 14, and both right parts of them are different from Fig. 11. If my comment is right and acceptable, the MA-A0E8 glyph for U+3C18 in UCIM should be as below. Note that the top left part has not been correct totally for Macao conventions.



Fig. 15 Suggested MA-A0E8 glyph in UCIM

Suggestions:

For Macao SAR,

- 1) Map MB1-C255 to U+5F5D (𢮑) which the glyph is as Fig. 11 shows and treat it as the suitable form. There is no need to do the horizontal extension now.
- 2) Add MD-5F5E for U+5F5E (𢮑) to MSCS which the glyph is as Fig. 12 shows and do the horizontal extension for the legacy use.
- 3) Add MA-A0E8 for U+3C18 (𢮑) to MSCS which the glyph is basically as Fig. 15 shows and do the horizontal extension for the IVS basic character use. Add ME-3C18-001 to MSCS which the glyph is as Fig. 13 shows. Add MA-A0E8 and ME-3C18-001 as MA_A0E8 and ME_3C18_001 to MSARG IVS collection for IVD.

For *TCA and CMEX*,

- 1) Register the Big5-2003, Big5+ and Big5 Eten Extended as a new “character set” to IANA, which is different from current Big5 (MIBenum: 2026) and [Big5HKSCS](#) (MIBenum: 2101). Maybe the better name is Big5-2003.
- 2) Ask the vendor to support for some necessary uses.
- 3) Provide a complete mapping table for Unihan.

For *UTC and ICU*,

- 1) Add a new property for the new “character set”.
- 2) Update the corresponding information in ICU.

For *Adobe* if possible,

- 1) Include all the glyphs for all the characters and IVSes used in Taiwan Province, Hong Kong SAR and Macao SAR to the future version of Adobe-CNS1.
- 2) Update the mappings with new characters and IVSes correspondingly.

2. MA-87DC

MA-87DC for U+2ADFF (灑) is inherited from HKSCS-2008 in UCIM, which is also only one H-Source character in CJK Ext. C, and the right part is related to MA-90F6 for U+9834 (穎). When we use the reference font in Table 1 to display these two characters, we will get the following results, which the glyph is not consistent.

Table 2 Current glyphs for U+9834 and U+2ADFF to follow the rules mentioned in Table 1

U+9834 MA-90F6	U+2ADFF MA-87DC

It is better to modify the glyph for U+2ADFF (灑) as the proper MA-87DC glyph as below.

Fig. 16 Suggested MA-87DC glyph in UCIM

If MSARG considers that it's OK to add glyph for U+2ADFF in PMingLiU-ExtB to MSCS, it is acceptable for IVS.

(End of Document)